PATENT ABSTRACTS OF JAPAN

(11)Publication number : 08-050499
(43)Date of publication of application : 20.02.1996

(51)Int.Cl. G10L 9/10
G06F 15/18
G10L 3/00
G10L 9/18

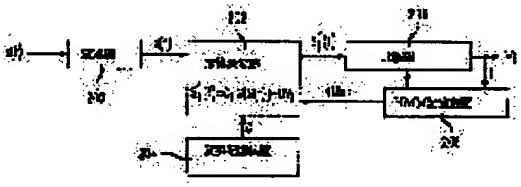
(21)Application number : 07-176872
(22)Date of filing : 21.06.1995

(71)Applicant : AT & T CORP
(72)Inventor : LEE CHIN-HUI
SANKAR ANANTH

(30)Priority
Priority number : 94 263284 Priority date : 21.06.1994 Priority country : US

(54) METHOD FOR DISCRIMINATING SIGNAL

(57)Abstract:
PURPOSE: To attain sound signal recognition suitable for application in an adverse environment by sharply reducing the average error rate of sound signal recognition under a non-coincident condition.
CONSTITUTION: A set of signals (e.g. a sound signal) and stored expressions (e.g. the stored expression of a keyword) is found out and at least one conversion for generating a signal for relatively and effectively emulating the stored expression is executed (202, etc.). The operation is attained by using one of three technical elements. In 1st technique, a signal is converted so as to be more approximated (e.g. approached) by one of plural stored expressions. In 2nd technique, a set of stored expressions is converted so that a signal is more approximated by one of the stored expressions. In 3rd technique, both of the signal and the set of stored expressions are converted. In this case, HMM, a neural network and a vector quantization expression are used as stored expressions.



LEGAL STATUS

[Date of request for examination]	05.03.1998
[Date of sending the examiner's decision of rejection]	
[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]	
[Date of final disposal for application]	
[Patent number]	3457431
[Date of registration]	01.08.2003
[Number of appeal against examiner's decision of rejection]	
[Date of requesting appeal against examiner's decision	

of rejection]
[Date of extinction of right]

(19)日本国特許庁（ＪＰ）　(12)公開特許公報（Ａ）　(11)特許出願公開番号

特開平8－50499

(43)公開日　平成８年(1996)２月20日

(51)Int.Cl.[6]		識別記号	庁内整理番号	ＦＩ	技術表示箇所
Ｇ１０Ｌ	9/10	３０１　Ｃ			
Ｇ０６Ｆ	15/18	５６０　Ｇ	8837－5Ｌ		
Ｇ１０Ｌ	3/00	５３５			
	9/18	Ｅ			

審査請求　未請求　請求項の数24　ＦＤ　（全 19 頁）

(21)出願番号　特願平7－176872

(22)出願日　平成７年(1995)６月21日

(31)優先権主張番号　２６３２８４

(32)優先日　1994年６月21日

(33)優先権主張国　米国（ＵＳ）

(71)出願人　390035493
エイ・ティ・アンド・ティ・コーポレーション
ＡＴ＆Ｔ　ＣＯＲＰ．
アメリカ合衆国　10013－2412　ニューヨーク　ニューヨーク　アヴェニュー　オブ　ジ　アメリカズ　32

(72)発明者　チンーファイ　リー
アメリカ合衆国，07974　ニュージャージー，ニュー　プロヴィデンス，ランニィメイド　パークウェイ　118

(74)代理人　弁理士　三俣　弘文

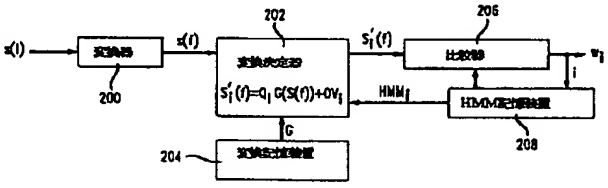
最終頁に続く

(54)【発明の名称】　信号識別方法

(57)【要約】

【目的】　不一致条件下で音声信号認識の平均誤り率を大幅に縮小し、悪環境での使用に適した音声信号認識を実現する。

【構成】　信号（例えば音声信号）と記憶された表現（例えばキーワードの記憶表現）の集合をとり、相対的に、記憶表現をより良好にエミュレートする信号を生じる少なくとも１回の変換を実行する。例えば、これは、３つの技術のうちの１つを使用することによって実現される。第１に、信号が記憶表現のうちの１つによってより良く近似される（例えば接近する）ように信号を変換することがある。第２に、記憶表現のうちの１つが信号をより良く近似するように記憶表現の集合を変換することがある。第３に、信号および記憶表現の集合の両方を変換することがある。記憶表現として、ＨＭＭ、ニューラルネットワーク、ベクトル量子化表現を用いることができる。



1

【特許請求の範囲】

【請求項１】　（ａ）信号を、記憶表現の集合と比較して、第１の類似度値の集合を生成するステップと、

（ｂ）前記信号の関数を計算するステップと、

（ｃ）前記信号と、前記関数の少なくとも一部と、前記第１の類似度値の集合内の少なくとも１つの類似度値に対応する少なくとも１つの記憶表現とに基づいて変換を決定するステップと、

（ｄ）前記変換によって前記信号を変換して変換済み信号を生成するステップと、

（ｅ）前記変換済み信号を前記記憶表現の集合と比較して第２の類似度値の集合を生成するステップと、

（ｆ）前記第２の類似度値の集合に基づいて、前記信号を、特定の記憶表現に類似するものとして識別するステップとからなることを特徴とする信号識別方法。

【請求項２】　前記記憶表現の集合内の各記憶表現は対応する動作を有し、前記特定の記憶表現に対応する動作を実行するステップをさらに有することを特徴とする請求項１の方法。

【請求項３】　前記信号が音声信号からなることを特徴とする請求項１の方法。

【請求項４】　（ａ）前記信号と、前記関数の少なくとも一部と、前記特定の記憶表現とに基づいて新たな変換を決定するステップと、

（ｂ）前記変換によって前記信号を変換して新たな変換済み信号を生成するステップと、

（ｃ）前記新たな変換済み信号を前記記憶表現の集合と比較して新たな類似度値の集合を生成するステップと、

（ｄ）前記新たな類似度値の集合内の少なくとも１つの値がしきい値より小さくなるまでステップａないしｃを反復するステップと、

（ｅ）前記信号を、前記新たな類似度値の集合内の少なくとも１つの値に対応する新たな特定の記憶表現に類似するものとして識別するステップとをさらに有することを特徴とする請求項１の方法。

【請求項５】　前記信号が周波数領域の信号からなることを特徴とする請求項１の方法。

【請求項６】　前記記憶表現の集合が隠れマルコフモデルの集合からなり、前記第１の類似度値の集合および前記第２の類似度値の集合が尤度値からなることを特徴とする請求項１の方法。

【請求項７】　前記記憶表現の集合がニューラルネットワークの集合からなり、前記第１の類似度値の集合および前記第２の類似度値の集合がニューラルネットワーク出力値からなることを特徴とする請求項１の方法。

【請求項８】　前記記憶表現の集合がベクトル量子化表現の集合からなり、前記第１の類似度値の集合および前記第２の類似度値の集合が歪み値からなることを特徴とする請求項１の方法。

【請求項９】　（ａ）信号を、記憶表現の集合と比較して、第１の類似度値の集合を生成するステップと、

2

（ｂ）前記信号と、前記第１の類似度値の集合内の少なくとも１つの類似度値に対応する少なくとも１つの記憶表現とに基づいて変換を決定するステップと、

（ｃ）前記変換によって前記記憶表現の集合を変換して変換済み表現の集合を生成するステップと、

（ｄ）前記信号を前記変換済み表現の集合と比較して第２の類似度値の集合を生成するステップと、

（ｅ）前記第２の類似度値の集合に基づいて、前記信号を、特定の記憶表現に類似するものとして識別するステップとからなることを特徴とする信号識別方法。

【請求項１０】　前記記憶表現の集合内の各記憶表現は対応する動作を有し、前記特定の記憶表現に対応する動作を実行するステップをさらに有することを特徴とする請求項９の方法。

【請求項１１】　前記信号が音声信号からなることを特徴とする請求項９の方法。

【請求項１２】　（ａ）前記信号と、前記特定の記憶表現とに基づいて新たな変換を決定するステップと、

（ｂ）前記変換によって前記記憶表現の集合を変換して新たな記憶表現の集合を生成するステップと、

（ｃ）前記新たな記憶表現の集合を前記信号と比較して新たな類似度値の集合を生成するステップと、

（ｄ）前記新たな類似度値の集合内の少なくとも１つの値がしきい値より小さくなるまでステップａないしｃを反復するステップと、

（ｅ）前記信号を、前記新たな類似度値の集合内の少なくとも１つの値に対応する新たな特定の記憶表現に類似するものとして識別するステップとをさらに有することを特徴とする請求項９の方法。

【請求項１３】　前記信号が周波数領域の信号からなることを特徴とする請求項９の方法。

【請求項１４】　前記記憶表現の集合が隠れマルコフモデルの集合からなり、前記第１の類似度値の集合および前記第２の類似度値の集合が尤度値からなることを特徴とする請求項９の方法。

【請求項１５】　前記記憶表現の集合がニューラルネットワークの集合からなり、前記第１の類似度値の集合および前記第２の類似度値の集合がニューラルネットワーク出力値からなることを特徴とする請求項９の方法。

【請求項１６】　前記記憶表現の集合がベクトル量子化表現の集合からなり、前記第１の類似度値の集合および前記第２の類似度値の集合が歪み値からなることを特徴とする請求項９の方法。

【請求項１７】　（ａ）信号と記憶表現の集合とを繰り返し変換して、信号を、現在の表現の集合内の少なくとも１つの記憶表現に近づけるステップと、

（ｂ）前記信号を、特定の記憶表現に類似するものとして識別するステップとからなることを特徴とする信号識別方法。

3

【請求項１８】　前記記憶表現の集合内の各記憶表現は対応する動作を有し、前記特定の記憶表現に対応する動作を実行するステップをさらに有することを特徴とする請求項１７の方法。

【請求項１９】　前記信号が音声信号からなることを特徴とする請求項１７の方法。

【請求項２０】　（ａ）前記少なくとも１つの記憶表現に対応する類似度値がしきい値より小さくなるまでステップａを実行することを特徴とする請求項１７の方法。

【請求項２１】　前記信号が周波数領域の信号からなることを特徴とする請求項１７の方法。

【請求項２２】　前記記憶表現の集合が隠れマルコフモデルの集合からなり、前記第１の類似度値の集合および前記第２の類似度値の集合が尤度値からなることを特徴とする請求項１７の方法。

【請求項２３】　前記記憶表現の集合がニューラルネットワークの集合からなり、前記第１の類似度値の集合および前記第２の類似度値の集合がニューラルネットワーク出力値からなることを特徴とする請求項１７の方法。

【請求項２４】　前記記憶表現の集合がベクトル量子化表現の集合からなり、前記第１の類似度値の集合および前記第２の類似度値の集合が歪み値からなることを特徴とする請求項１７の方法。

【発明の詳細な説明】

【０００１】

【産業上の利用分野】本発明は、信号認識に関し、特に、悪環境での使用に適した自動信号認識システムの性能を改善することに関する。

【０００２】

【従来の技術】信号認識システムの１つのタイプは音声認識システムである。

【０００３】音声認識システム（以下単に「システム」という。）は、話者から受け取った入力（すなわち音声）において単語の集合を認識することができる。認識した入力に基づいて、機能が実行される。システムを電話網で使用する場合、その機能は例えば話者をオペレータに接続することである。

【０００４】システムは、単語の集合内の各単語を認識するように事前にトレーニングされる。各単語をキーワードという。トレーニングは、トレーニング音声を入力し、キーワードのモデルを形成して記憶することによって行うことができる。

【０００５】トレーニングされると、動作時には、システムは入力音声内に含まれるキーワードを認識することができる。システムは、入力音声を記憶されたキーワードモデルと比較することによってこれを行う。入力音声内のキーワードが認識されると、システムはその単語に関係づけられた機能を実行する。

【０００６】既知のシステムはかなり高いレベルの精度で入力音声内のキーワードを識別するが、多くの改良の

4

余地がある。システムの精度は、システムの「平均単語誤り率」によって測定することができる。「平均単語誤り率」は、システムがキーワードを含む発声内で誤ったキーワードを認識するかまたはキーワードが発声されていないときにキーワードを識別する頻度の測度である。

【０００７】システムの精度を低下させる１つの要因は「不一致」である。不一致は、システムがある入力システム（例えば、マイクロホンおよびケーブル）を使用してトレーニングされ、別の入力システム（例えば、電話ハンドセットおよび接続された電話網システム）で使用されるときに生じることがある。不一致は、この例の場合、人の声がマイクロホンを通過するときに電話システムの場合と異なる特性を示すために生じるといわれている。

【０００８】精度の問題、特に不一致の問題は、これまで注意を受けている。この不一致問題を解決しようとした少なくとも３つの方法がある。

【０００９】第１に、話者が入力音声を供給する方法に依存してアプリケーションごとにシステムを訓練することによって、不一致問題を解決することが提案されている。この提案は２つの問題点を有する。第１に、アプリケーションごとにシステムをトレーニングすることは時間を消費し高価である。第２に、システムは、異なる音声入力媒体（例えば、電話およびセルラ電話）がシステムで使用可能である場合には、やはり不一致問題の影響を示すことになる。

【００１０】第２に、システムで使用する「プール」モデルを作成することによって不一致問題を解決することが提案されている。複数の入力媒体の効果の混合物を反映するモデルがある。この提案にも２つの問題点がある。第１に、プールモデルは、他のモデルよりも作成するのが高価である（そして一般に時間もかかる）。第２に、与えられた任意の入力システムに対して、その入力システムに基づいて作成されるモデルはプールモデルよりも高い精度を有する。

【００１１】第３に、システムに入力される音声信号にオフセット因子を付加することによって不一致問題を解決することが提案されている。付加されるオフセット因子は、信号が受ける歪みを相殺するのに必要なものの推定値を表す。このオフセット因子はコードブック探索によって決定される。コードブックは多くのエントリ（代表的には２５６個）からなる。これらのエントリは、入力システムの集合からのオフセット値を表す。例えば、コードブックは、第１のタイプの入力システム（例えば、第１のタイプのマイクロホンおよび第１のタイプのチャネル）に基づく６４個のエントリと、第２のタイプの入力システムに基づく第２の６４個のエントリと、第３のタイプの入力システムに基づく第３の６４個のエントリと、第４のタイプの入力システムに基づく第４の６４個のエントリとを有する。しかし、この提案にも問題

点がある。音声信号として使用される入力が第１、第２、第３、または第４のタイプの入力システムではない場合には、使用される入力が２５６個のコードブックエントリの集合によって正しく特徴づけられないことがある。これによって、音声認識システムの精度は悪くなる。

【００１２】

【発明が解決しようとする課題】上記の方法の使用にもかかわらず、不一致条件下での認識に対する平均単語誤り率は多くのアプリケーションでは未だに不満足なものである。不一致問題の解決が必要とされる。

【００１３】

【課題を解決するための手段】本発明は、不一致条件下で信号の平均誤り率を大幅に縮小する方法を提供する。本発明の方法は、信号（例えば音声信号）と記憶された表現（例えばキーワードの記憶表現）の集合をとり、相対的に、記憶表現をより良好にエミュレートする信号を生じる少なくとも１回の変換を実行する。例えば、これは、３つの技術のうちの１つを使用することによって実現される。第１に、信号が記憶表現のうちの１つによってより良く近似される（例えば接近する）ように信号を変換することがある。第２に、記憶表現のうちの１つが信号をより良く近似するように記憶表現の集合を変換することがある。第３に、信号および記憶表現の集合の両方を変換することがある。

【００１４】本発明に従って形成したシステムは、個々のアプリケーションごとにトレーニングする必要はないという効果がある。

【００１５】また、本発明に従って形成したシステムは、プールモデルやコードブックを使用する既知のシステムよりも高い精度を有するという効果がある。

【００１６】

【実施例】ここでは、音声認識システムで本発明を実施するという状況で説明する。しかし、当業者には明らかなように、本発明は、物理システムからの物理信号が認識のために記憶表現の集合と比較されるような任意のタイプのシステム（例えば、ファクシミリシステム、衛星システム、光学式文字認識システムなど）で使用可能である。さらに、「強固な音声認識のための確率論的マッチングへの最尤アプローチ(A Maximum Likelihood Approach to Stochastic Matching for Robust Speech Recognition)」という題名の論文を付録Ａとして記載する。付録Ａは本発明を例示することのみを意図したものである。

【００１７】説明を明確にするため、本発明の実施例は個別の機能ブロックからなるものとして表す。これらのブロックが表現する機能は、共用または専用のハードウェアの使用によって実現される。ハードウェアには、ソフトウェアを実行可能なハードウェアも含まれるが、それに限定されるものではない。例えば、図２のブロック

２００、２０２、および２０６の機能は単一の共用プロセッサによって実現することができる。「プロセッサ」という用語の使用は、ソフトウェアを実行可能なハードウェアのみを指すと解釈してはならない。

【００１８】実施例は、ＡＴ＆ＴのＤＳＰ１６またはＤＳＰ３２Ｃのようなディジタル信号プロセッサ（ＤＳＰ）と、以下で説明する動作を実行するソフトウェアを記憶する読み出し専用メモリ（ＲＯＭ）と、ＤＳＰの結果を記憶するランダムアクセスメモリ（ＲＡＭ）とからなることが可能である。超大規模集積（ＶＬＳＩ）ハードウェア実施例や、カスタムＶＬＳＩ回路と汎用ＤＳＰ回路の組合せも実現可能である。

【００１９】音声認識システムの状況で本発明を説明する前に、既知の音声認識システムについて簡単に説明する。

【００２０】図１を参照すると、音声信号Ｓ（ｔ）が変換器１００に入力される。簡単のため、Ｓ（ｔ）は、１個のキーワードを含みそれ以外は含まない発声であると仮定する。変換器１００は音声信号Ｓ（ｔ）を周波数領域に変換し、信号Ｓ（ｆ）を生成する。Ｓ（ｆ）はオフセット計算器１０２に入力される。オフセット計算器１０２はS_i'（ｆ）を出力する。S_i'（ｆ）はＳ（ｆ）にコードブック１０４に記憶されているオフセット値OV_iを加えたものに等しい。S_i'（ｆ）は比較器１０６に入力され、そこで、HMM記憶装置１０８に記憶されている隠れマルコフモデル（HMM）の集合と比較される。HMM記憶装置１０８内の各HMMはキーワードに対するHMMである。S_i'（ｆ）を各HMMと比較した後、比較器１０６は、最も良くS_i'（ｆ）に一致するHMMによって表現される単語を出力する。この単語をW_iで表す。次に、このプロセスはコードブック１０４内の各オフセット値に対して反復される。各OV_iに対して、新たなW_iが比較器１０６によって決定される。W_iが各オフセット値OV_iに対して決定されると、最も高い尤度スコアのW_iが、発声に含まれるキーワードを表すものとして選択される。

【００２１】既知のシステムについて説明したので、次に、本発明を、３つの異なる実施例における音声認識システムについて説明する。第１実施例は、音声信号の表現（例えばケプストラム表現）を、記憶されている少なくとも１つのHMMにより良く類似するように変換する。第２実施例は、記憶HMMを、音声信号の表現により良く類似するように変換する。第３実施例は、信号および記憶表現の集合の両方を変換し、信号が記憶HMMに類似しているよりも、変換した信号が変換した少なくとも１つのHMMに良く類似するようにする。

【００２２】図２を参照して第１実施例について説明する。音声信号Ｓ（ｔ）は変換器２００に入力される。音声認識環境では、変換器２００は任意のタイプの特徴抽出装置（例えば、スペクトラム分析装置、ケプストラム

分析装置など）である。再び簡単のため、S（t）は1個のキーワードを含む発声であると仮定する。変換器200は音声信号S（t）を周波数領域に変換し信号S（f）を生成する。S（f）は変換プロセッサ202に入力される。これは、図2および図3に示されているように変換決定器とも呼ぶ。変換プロセッサ202はSi′（f）を出力する。Si′（f）はQiG（S（f））にオフセット値OViを加えたものに等しい。Qiはスケールファクタであり、Gは、変換記憶装置204に記憶されているS（f）の関数である。S（f）の「関数」は、変換プロセッサで計算される。この関数は必ずしもQiを含むとは限らない。しかし、この「関数」は、少なくとも3つのことなる形式を含むほどに広いものである。特に、この「関数」はQiG（S（f））、QiS（f）、またはG（S（f））の形式が可能である。Gは任意の線形または非線形の関数であり、その形式はアプリケーションに依存し、信号タイプ、予想される歪み、および特徴抽出方法（例えば、ケプストラム、スペクトラムなど）を含む因子に依存する。Gが上記の因子に基づいて変化するしかたは当業者には容易に明らかとなる。例えば、信号タイプが音声であり、予想される歪みが電話網のチャネル歪みであり、特徴抽出方法がケプストラム分析である場合、G（x）＝xである。

【0023】再び図2を参照すると、変数iを識別するために、システムの初期通過（初期識別）が必要である。初期通過中にはSi′（f）はS（f）に等しく（すなわち、Qi＝1、G（x）＝x、およびOVi＝0）、比較器206は単にS（f）と各HMMの間の比較を実行して第1の尤度値の集合を決定する。この第1尤度値集合に基づいて、Wiを決定する。Wiは、初期通過でのSi′（f）に最も近いHMM（すなわちHMMi）によって表現される単語である。

【0024】さらに図2を参照すると、初期識別後、変換プロセッサ202はHMM記憶装置208からHMMiを受信する。HMM記憶装置208は、初期識別後に比較器206からインデックスiを送信されるため、HMMiを選択する。変換プロセッサ202はS（f）、HMMi、および変換記憶装置からの入力Gを使用して、関数を計算し、スケールファクタQiおよびオフセット値OViを決定する。QiおよびOViは付録Aの式32および式33に従って決定される。OVi、QiおよびGはこれで既知となるため、変換プロセッサは、S（f）をとり、上記の式に従ってこれを新たなSi′（f）に変換することが可能となる。

【0025】再び図2を参照すると、Gは数学的にQiG（S（f））と記述されるS（f）の関数の一部である。変換（例えばSi′（f）＝Qi（G（S（f））＋OVi）が変換プロセッサ202によって決定されるときに「決定される」必要がある項目はQiおよびOViであるため、これは重要である。方程式が2個および未知数が2個あるため、QiおよびOViはどちらを先に（すなわちQiが先でOViが後またはOViが先でその後にQi）決定することも可能である。この変換を決定することはある状況では関数全体（例えばQiおよびG）ではなく関数の一部（例えばG）に基づくこともある。例えば、関数がQiG（S（f））であるとみなされるとする。この関数は計算可能である。しかし、OViについて先に解きQiについて後に解くことによって変換を決定する場合、関数の少なくとも一部（すなわちQi）は変換のOVi部分を決定するのに必要ではない。OViが決定されると、Qiを決定することは自明でありOViに依存する。従って、変換は関数の少なくとも一部に基づいて決定されることがある。

【0026】再び図2を参照すると、新たなSi′（f）を各HMMと比較し、第2の尤度値の集合を決定する。この第2尤度値集合に基づいてWiを決定する。この新たなWiは新たなSi′（f）に最も近いHMMである新たなHMMiに対応する。このプロセスは反復されることもあるが、アプリケーションによっては反復する必要はない。例えば、新たなWiからのインデックスiは次のHMMiを識別するHMM記憶装置208に渡される。次のHMMiは、S（f）およびGとともに、変換プロセッサに入力される。この情報により、変換プロセッサは次のQiおよび次のOViを決定する。

【0027】ある場合には、第2尤度値集合に基づいて、新たなSi′（f）が特定のHMMに最も類似するものとして識別される。しかし、図2について説明したプロセスは反復されることが可能であり、プロセスの追加反復が所望されるような場合もある。このような場合には、プロセスの反復を何回実行するかを決定するいくつもの方法がある。例えば、特定のHMMiが現在のSi′（f）に対するあるしきい値より高くなった場合に追加反復を行わないと決定することも可能である。また、ある一定回数の反復より多くを行わないと決定することも可能である。また、以上の2つの方法を組み合わせて、あるしきい値より高いHMMiがない限り一定回数の反復を行うことも可能である。このほかにも、当業者には容易に明らかなように、プロセスの反復を何回実行するかを決定するために使用可能な他のいくつもの方法がある。

【0028】図3を参照して第2実施例について説明する。音声信号S（t）は変換器300に入力され、音声信号の周波数表現S（f）を生成する。再び簡単のため、S（t）は1個のキーワードを含む発声であると仮定する。S（f）は比較器302に入力される。システムの初期通過中には、S（f）はHMMの集合（例えばHMM′）と比較される。HMM′は、HMM記憶装置304に記憶されているHMMの別の集合に基づいたHMMの集合である。HMM′内の各HMMはHMM記憶

装置のHMMにオフセット値ＯＶを加えたものに等しい。初期通過の場合、ＯＶは０に等しく設定される。初期通過は第１の尤度値の集合を生成する。HMM′内で最も高い尤度値を与える特定のHMMによって表される単語Wiが識別される。このHMMは、例えば、HMM′内のｉ番目のHMMである。こうして、Wiは、音声信号Ｓ（ｔ）に最も近い単語の初期決定を表す。

【００２９】Wiのインデックス（すなわちｉ）はHMM記憶装置３０４に入力される。HMM記憶装置３０４はこのｉ番目のHMM（すなわちHMMi）を識別し、それを変換プロセッサ３０６に入力する。HMMｉおよびＳ（ｆ）に基づいて、変換プロセッサ３０６はオフセット値ＯＶを決定する。ＯＶは、HMMi、Ｓ（ｆ）、および付録Ａの式４７および式４９に基づいて決定される。ＯＶは、HMMi、Ｓ（ｆ）、および付録Ａの式５３および式５４に基づいて決定することも可能である。一般に、式４７および式４９の使用は、式５３および式５４の使用よりも計算量的にわずかに効率的であるがわずかに有効でない。

【００３０】オフセット値をHMM記憶装置の各HMMに加えて新たなHMMの集合（すなわち新たなHMM′）を生成する。これは変換プロセッサ３０６で行われる。こうして、HMM記憶装置３０４内の記憶HMMの集合が、変換されたHMMの集合（すなわち新たなHMM′）を生成するために変換される。

【００３１】この新たなHMM′は比較器３０２でＳ（ｆ）と比較される。この比較により、第２の尤度値の集合が生成される。この第２尤度値集合に基づいてWiが決定される。この新たなＷｉは、新たなHMM′内のHMMに対応する。このプロセスは反復されることもあるが、アプリケーションによっては反復する必要はない。

【００３２】プロセスの追加反復が所望されることもある。そのような場合、当業者には容易に明らかなように、プロセスの反復を何回実行するかを決定するいくつもの方法がある。

【００３３】第３実施例は、第１実施例および第２実施例の両方からの技術を使用する。従って、第３実施例は図２および図３を参照して説明する。

【００３４】第３実施例は、信号と、記憶HMMの集合内の少なくとも１つのHMMを互いに接近させる反復法を使用する。例えば、Ｓ（ｔ）が図２のシステムに入力されWiが生成される。しかし、インデックスｉを図２のHMM記憶装置２０８に送る代わりに、インデックスｉは図３のHMM記憶装置３０４に送られる。次に、図３のシステムはHMMiおよびＳ（ｔ）を使用して新たなWiを生成する。第３実施例に従って形成されるシステムによって実行されるプロセスは、１回だけが所望される場合には完了しない可能性もある。しかし、アプリケーションに基づいて追加反復が所望される場合、図３

の新たなWiは新たなインデックスｉを図３のHMM記憶装置３０４ではなく図２のHMM記憶装置２０８に送る。次に、図２のシステムはプロセスを実行して現在のWiを生成する。次に、現在のWiに対応するｉの値が再び図３のHMM記憶装置３０４に送られる。これは必要なだけ続けることができる。追加反復が所望される場合、当業者には容易に明らかなように、プロセスの反復を何回実行するかを決定するいくつもの方法がある。

【００３５】第３実施例を参照すると、２つの事項が当業者には明らかである。第１に、実行する反復回数にかかわらず、第３実施例は、図２または図３に関して説明した技術のいずれを最初に実行することによっても使用可能である。第２に、反復はさまざまな方法で実行可能である。例えば、次のような方法がある。

１．図２に関して説明した初期識別手順を実行した後、図３に関して説明した手順を実行するか、またはその逆。

２．図２に関して説明した初期識別手順およびその後の手順を実行した後、図３に関して説明した手順を実行するか、またはその逆。

他の可能性は当業者には明らかである。

【００３６】本発明は音声認識システムに関して説明したが、当業者には明らかなように、本発明は、物理信号を記憶された表現の集合と比較するような任意のタイプのシステムで使用可能である。そのようなシステムには、ファクシミリ認識システム、衛星伝送・認識システム、話者認識システム、署名認識システム、および光学式文字認識システムならびにその他の画像または音響認識システムがあるが、これらに限定されるものではない。さらに、各記憶表現は、その記憶表現が信号に類似するものとして識別された場合に実行する対応する動作を有することが可能である。この動作はアプリケーションに依存し、例えば、電話網では発呼者をオペレータに接続することである。当業者には理解されるように、アプリケーションに依存して実行されるさまざまな動作がある。最後に、開示したシステムは単語を表す記憶モデルに関して説明したが、当業者には理解されるように、この記憶モデルは、アプリケーションに依存して、音素、または、音声もしくはデータの他の要素のモデルとすることも可能である。

【００３７】［付録Ａ］

要約

試験発声における歪みによって引き起こされる認識性能劣化を縮小するために、試験発声と与えられた音声モデルの集合の間の音響不一致を減少させる最尤（ＭＬ）確率論的不一致アプローチを開示する。音声信号が部分語隠れマルコフモデル（HMM）の集合Λxによってモデル化されると仮定する。観測される試験発声Ｙとモデル Λxの間の不一致は２つの方法、すなわち、（１）Ｙを、モデルΛxとより良く一致する発声Ｘに写像する逆

11

歪み関数Ｆν（・）によって、および、（２）Λxを、発声Ｙにより良く一致する変換されたモデル（変換済みモデル）Λyに写像するモデル変換関数Ｇη（・）によって、縮小することができる。変換Ｆν（・）またはＧη（・）の関数形を仮定し、期待値最大化（ＥＭ）アルゴリズムを使用して最尤的にパラメータνまたはηを推定する。Ｆν（・）またはＧη（・）の形の選択は、音響不一致の性質の事前の知識に基づく。

【００３８】実験結果を提示して、提案するアルゴリズムの性質を調べ、異なるトランスデューサおよび伝送チャネルによる不一致の存在するＨＭＭベースの連続音声認識システムの性能の改善におけるこのアプローチの効力を確認する。提案する確率論的マッチングアルゴリズムは急速に収束することがわかる。さらに、不一致条件における認識性能は大幅に改善される一方、一致条件における性能も良好に維持される。不一致条件におけるこの確率論的マッチングアルゴリズムによる平均単語誤り率の縮小は約７０％である。

【００３９】１．はじめに

最近、悪環境での自動音声認識（ＡＳＲ）システムの性能を改善するという問題に関心が集まっている。トレーニング環境と試験環境の間に不一致があると、ＡＳＲシステムの性能は劣化する。強固な音声認識の目標は、この不一致の影響を除去して一致条件にできるだけ近い認識性能を実現することである。音声認識では、音声は通常隠れマルコフモデル（ＨＭＭ）Λxの集合によってモデル化される。認識中、観測される発声Ｙはこれらのモデルを使用して復号される。トレーニング条件と試験条件の間の不一致により、この性能は不一致条件に比較して劣化することが多い。

【００４０】トレーニング条件と試験条件の間の不一致は、図４に示した信号空間、特徴空間、またはモデル空間で見ることができる。図４では、Ｓはトレーニング環境における原音声を示す。トレーニング環境と試験環境の間の不一致は、ＳをＴに変換する歪みD_1によってモデル化される。音声認識では、まず何らかの形の特徴抽出が実行される。その特徴を、トレーニング環境ではＸで表し、試験環境ではＹで表す。特徴空間におけるこれら２つの環境の間の不一致は、特徴Ｘを特徴Ｙに変換する関数D_2によってモデル化される。最後に、特徴Ｘを使用してモデルΛxを構築する。モデル空間では、トレーニング環境と試験環境の間の不一致は、ΛxをΛyに写像する変換D_3とみなすことができる。不一致の原因には、加法的ノイズ、スペクトラム傾斜およびスペクトラム形成に寄与するチャネルとトランスデューサの不一致、話者の不一致、異なるアクセント、強制、および異なる話し方がある。最近の多くの研究は、加法的ノイズおよびチャネルの効果の問題に集中している。

【００４１】ノイズに対処するために使用される方法は一般的に３つの大まかなカテゴリーに分けられる。第１

12

のカテゴリーでは、強固な信号処理を使用して、可能な歪みに対する特徴の感度を減少させるものである。１つのアプローチでは、リフタリングのようなスペクトラム形成を行う。その考え方は、低次および高次のケプストラム成分をデエンファサイズすることである。これらの成分は、チャネルノイズおよび加法的ノイズの効果に敏感であることがわかっているからである。発声からの長時間ケプストラム平均を減算することに基づく方法も提案されている。この考え方は、チャネルによる不一致を除去するために一般的に使用される。またこの第１のカテゴリーには、スペクトラムシーケンスをハイパスフィルタリングして緩変動チャネルの効果を除去する方法もある。聴覚モデリングに基づく方法では、信号処理を使用して人間の耳の処理を模倣し、より強固な特徴が得られることを期待する。音声特徴へのノイズの影響を縮小することができる信号制限プリプロセッサの使用も知られている。ノイズの影響を縮小するもう１つの方法は、トレーニングデータに環境ノイズの一部を注入してシステムを再トレーニングすることである。この技術はディザリングに似ている。また、スペクトラム減算に基づく方法もある。この方法では、ノイズパワースペクトラムの推定値を各音声フレームから減算する。この第１のカテゴリーのアプローチは、一般に何らかの形の強固な特徴前処理を含むので、特徴空間（図４）で作用するとみなすことができる。

【００４２】第２のカテゴリーは、何らかの最適基準を使用して、明瞭な音声の関数の推定値を形成することである。音声スペクトラムの関数の最小平均２乗誤差（ＭＳＥ）推定に基づく定式化において、破壊的ノイズが独立のガウス過程であると仮定するものが知られている。さらに、各スペクトラムビンは別個に推定される。個々のビンは独立であると仮定したためである。音声分布をガウシアンの混合（ミクスチャ）としてモデル化して、別々のスペクトラムビン間の相関を各ミクスチャの対角共分散行列でモデル化する。最後に、音声の時間構造は、隠れマルコフモデル（ＨＭＭ）によって音声をモデル化することにより考慮される。これらのアプローチは、信号空間と特徴空間のいずれの表現を推定しているかに依存して、音声強化として信号空間で、または、スペクトラム補償として特徴空間で、見ることができる。

【００４３】第３のカテゴリーでは、ノイズをモデル化し直接認識プロセスに組み込む。このアプローチの１つは、ノイズマスキングに基づく。この考え方では、信号エネルギーがある適当なノイズレベル以下になった場合に、そのノイズレベルでフィルタバンクエネルギーを置き換える。こうして、ノイズによって顕著に破壊された情報は無視される。モデル分解と呼ばれるもう１つのアプローチでは、音声およびノイズの別々のＨＭＭをトレーニングデータからトレーニングする。認識中には、これらの２つのモデルの結合状態空間でビタビ検索を実行

する。この方法はかなり良好に動作することが示されているが、音声およびノイズの両方に対する精度の良いモデルが必要である。このモデル分解アプローチは上記のアプローチに似ている。しかし、原音声のパラメータは、認識中のノイズのある音声から推定される。信号とノイズのモデルの間のさらに一般的な相互作用を許容した、ノイズのあるデータから原音声パラメータを推定する問題は良く研究されている。この場合、信号はガウシアンのミクスチャとしてモデル化されると仮定される。このカテゴリーのさらにもう１つの方法では、HMMパラメータを推定する前に、トレーニング音声のエネルギー等高線の最尤（ＭＬ）推定を使用して音声を正規化する。試験中には、明瞭な利得パラメータのＭＬ推定がノイズのある音声から計算され、それを使用して、音声モデルのパラメータを正規化する。音声認識に対するミニマックス法は既知であり、認識器は、トレーニング中に推定された値の近傍をHMMパラメータが占有するようにすることによってさらに強固となる。これらのアプローチは、図４に示したモデルにおける可能な歪みを扱うモデル空間で作用するとみなされる。

【００４４】話者とチャネルの適応に関する最近の研究では、各話者を基準話者に変換する固定バイアスを推定した後、推定されたバイアスを各音声フレームから減算する。類似のアプローチが、音声がベクトル量子化（ＶＱ）コードブックによってモデル化されるような音声認識においてチャネル不一致を推定するために使用されている。チャネル不一致を推定するもう１つのアプローチとして、推定が、２つのチャネルの平均対数スペクトラム間の差に基づいて行うことが提案されている。

【００４５】ここに開示するのは、強固な音声認識のための確率論的マッチングへのＭＬアプローチである。本方法では、発声の認識中に、ＭＬアプローチを使用することによって、観測発声と原音声モデルの間の不一致を縮小する。この不一致は少なくとも２つの方法で縮小することができる。第１に、歪んだ特徴Ｙを原特徴Ｘの推定値に写像し、原モデルΛ_Xを認識に用いることができるようにすることが可能である。第２に、原モデルΛ_Xを、観測発声Ｙにより良く一致する変換済みモデルΛ_Yに写像することが可能である。第１の写像は特徴空間で作用し、第２の写像はモデル空間で作用する（図４）。これらの写像を、特徴空間ではF_ν（Ｙ）で表し、モデル空間ではG_η（Λ_X）で表す。ただしνおよびηは推定すべきパラメータである。これらの写像の関数形は、音響不一致の性質に関する事前の情報に依存する。次に、与えられたモデルΛ_Xに対して観測音声Ｙの尤度を最大化して歪みによる不一致を減少させるようにこれらの関数のパラメータνまたはηを推定する。目標は認識を改善するために不一致を縮小することであり、Λ_Xは認識に使用するモデルであるので、パラメータνおよびηを推定するための音声モデルとしてHMMのΛ_Xを使

用することは直感的に興味のあることである。ＭＬパラメータ推定は、反復して尤度を改善する期待値最大化（ＥＭ）アルゴリズムを使用して解かれる。本発明の確率論的マッチングアルゴリズムは与えられた試験発声および与えられた音声モデルの集合のみに作用するため、実際の試験前の不一致の推定にトレーニングは不要である。上記の２つの方法をともに使用して、不一致の効果を縮小する第３の方法とすることも可能である。

【００４６】異なるトランスデューサおよびチャネルによる不一致の存在下で連続音声認識システムの性能を改善するアプローチの効力を示すために実験結果を提示する。この不一致は、固定バイアス（特徴空間において）として、およびランダムバイアス（モデル空間において）としての両方でモデル化される。提案するアプローチは、不一致条件で単語誤り率を約７０％縮小し、一致条件下での性能を維持した。本発明のアルゴリズムは、急速に収束する（反復２回以内）こともわかった。

【００４７】以下の説明の構成は次の通りである。第２節で、変換F_ν（・）およびG_η（・）のパラメータの最尤推定に対する一般的な枠組みを説明する。第３節で、特徴空間変換の場合を説明する。特に、未知パラメータに関して線形であるが、観測値に関して非線形な逆歪み関数のパラメータの推定値に対する表式を導出する。特別な場合として、加法的バイアスモデルを考察する。第４節で、変換をモデル空間で見る。特に、ランダム加法的バイアス歪みの場合を考察する。

【００４８】２．確率論的マッチングの構成

パターン認識では、トレーニングモデルの集合$\Lambda_X = \{\lambda_{Xi}\}$（ただし、$\lambda_{Xi}$はｉ番目のクラスのモデルである。）、および試験データの集合$Y = \{y_1, y_2, \ldots, y_T\}$が与えられた場合に、Ｙに埋め込まれた事象の列$W = \{W_1, W_2, \ldots, W_L\}$を認識することが所望される。連続音声認識の場合、例えば、λ_{Xi}はｉ番目の部分語HMM単位に対応し、Ｙは特定の試験発声に対応する。その場合Ｗは復号された単音または単語の列となる。モデルΛ_Xをトレーニングする際には、トレーニングデータの集合に制限される。残念ながら、このトレーニングデータと試験データＹの間に不一致が存在することがあり、このことは、認識される列Ｗにおける誤りを引き起こす。この不一致は、もとの信号空間、特徴空間、またはモデル空間（図４）で見ることができる。図中、関数Ｄ（・）は原空間を対応する歪みのある空間に写像する。不一致の原因には、信号中の歪み、信号の不完全な特徴づけ、不十分な量のトレーニングデータ、または不適切なモデル化および推定誤差がある。以下では、トレーニング音声データと試験音声データの不一致による音声認識性能の劣化の問題を考える。この不一致は、マイクロホンとチャネルの不一致、トレーニングと試験の環境の相違、話者および話し方またはアクセントの相違、またはこれらの任意の組合せによる可能性

15

がある。

【００４９】音声認識では、モデルΛxを使用して、最 ＊

$$W' = \arg\max_{W} p(Y|W,\Lambda_X)P(W)$$

を用いてＹを復号する。ただし、第１項は単語列Ｗが与えられた場合にＹを観測する尤度であり、第２項は単語列Ｗの事前確率である。この第２項は、許容単語列の集合に制約を加える言語モデルと呼ばれる。トレーニング環境と試験環境の間の不一致により、式１によって評価されるΛxが与えられたときのＹの尤度に対応する不一 10
致があり、復号された列Ｗ′に誤りを引き起こす。この※

$$X = F_{\nu}(Y)$$

ただし、νは逆歪み関数のパラメータである。あるいは、モデル空間において、パラメータηを有しΛxを変 ★

$$\Lambda_Y = G_{\eta}(\Lambda_X)$$

ＹとΛxの間の不一致を減少させる１つのアプローチは、モデルΛxが与えられたときの式１のＹとＷの結合尤度を最大にするパラメータνまたはηおよび単語列Ｗ☆ 20

$$(\nu',W') = \arg\max_{(\nu,W)} p(Y,W|\nu,\Lambda_X)$$

これは次式と等価である。

$$(\nu',W') = \arg\max_{(\nu,W)} p(Y|W,\nu,\Lambda_X)P(W)$$

対応して、モデル空間では、次のようなη′を見つけることが必要である。 ◆

$$(\eta,W') = \arg\max_{(\eta,W)} p(Y,W|\eta,\Lambda_X)$$

これは次式と等価である。

$$(\eta',W') = \arg\max_{(\eta,W)} p(Y|W,\eta,\Lambda_X)P(W)$$

式５における変数νとＷに関する、または、式７におけるηとＷに関するこの同時最大化は、νまたはηを固定してＷについて最大化し、その後、Ｗを固定してνまたはηについて最大化することを反復することにより行わ 40
れる。この手続きを、概念的に、特徴空間について図５に、また、モデル空間について図６に示す。 ＊

$$\nu' = \arg\max_{\nu} p(Y|\nu,\Lambda_X)$$

および

$$\eta' = \arg\max_{\eta} p(Y|\eta,\Lambda_X)$$

のように書く。

【００５２】これを調べるため、Λxが左右連続密度部 50

16

＊大事後（ＭＡＰ）デコーダ

【数１】

(1)

※不一致を減少させることにより認識性能が改善される。

【００５０】特徴空間において、歪み関数が原発声Ｘ＝｛x1, x2, ..., xT｝を観測値の列Ｙ＝｛y1, y2, ..., yT｝に写像するとする。この歪みが可逆である場合、次のような逆関数Ｆνでを原音声Ｘに写像することができる。

【数２】

(2)

★換済みモデルΛＹに写像する変換Ｇηを考える。

【数３】

(3)

☆を見つけることである。すなわち、特徴空間では、次のようなν′を見つけることが必要である。

【数４】

(4)

【数５】

(5)

◆【数６】

(6)

【数７】

(7)

＊【００５１】Ｗを見つけるプロセスは多くの研究者によって扱われている。パラメータνおよびηを見つける問題は興味がある。表式を簡単にするため、Ｗ依存性を除去し、式５および７に対応する最尤推定問題を

【数８】

(8)

【数９】

(9)

分語ＨＭＭの集合であると仮定する。ｉ，ｊ＝１，...，Ｎに対して、状態ｉからｊへの遷移確率をａ

17

i, jで表し、状態 i に対する観測値密度 p x （ x | i ）は次式によって与えられるようなガウシアンのミクスチャ＊

$$p_x(x|i)=\sum_{j=1}^{M} w_{i,j}\,N(x;\mu_{i,j},C_{i,j}) \qquad (10)$$

ただしMはミクスチャの数であり、wi, jは状態 i におけるミクスチャ j の確率であり、Nは次式によって与え※

$$N(x;\mu_{i,j},C_{i,j})=\frac{1}{(2\pi)^{D/2}|C_{i,j}|^{1/2}}\exp\left(-\frac{1}{2}(x-\mu_{i,j})^{T}C_{i,j}^{-1}(x-\mu_{i,j})\right) \qquad (11)$$

ただしDは特徴ベクトルxの次元であり、Ci, jおよびμi, jは状態 i におけるミクスチャ j に対応する共分散行列および平均値ベクトルである。

【００５３】Ｓ＝｛ｓ１，ｓ２，...，ｓＴ｝を、モデルの集合ΛＸに対するすべての可能な状態列の集合とし、Ｃ★

$$\nu'=\arg\max_{\nu} p(Y|\nu,\Lambda_X)=\arg\max_{\nu}\sum_{S}\sum_{C} p(Y,S,C|\nu,\Lambda_X) \qquad (12)$$

同様に、式９は次のように書くことができる。

$$\eta'=\arg\max_{\eta}\sum_{S}\sum_{C} p(Y,S,C|\eta,\Lambda_X) \qquad (13)$$

【００５４】一般に、ν′またはη′を直接推定することは容易ではない。しかし、あるＦνおよびＧηに対しては、ＥＭアルゴリズムを使用して現在の推定値を反復的に改善し、式１２および１３中の尤度が反復ごとに増大するように新たな推定値を得ることができる。次の２つの節で、特徴空間変換Ｆνのパラメータν、およびモデル空間変換Ｇηのパラメータηの推定値を見つけるためのＥＭアルゴリズムの応用をそれぞれ説明する。

☆【００５５】３．特徴空間変換Ｆνの推定

この節では、ＥＭアルゴリズムを使用して式８の推定値ν′を見つける。ＥＭアルゴリズムは２ステップ反復手続きである。第１ステップは、期待値ステップ（Ｅステップ）と呼ぶ。この第１ステップで、次式によって与えられる補助関数を計算する。

$$Q(\nu'|\nu)=E\left\{\log p(Y,S,C|\nu',\Lambda_X)\,|\,Y,\nu,\Lambda_X\right\} \qquad (14)$$
$$=\sum_{S}\sum_{C} p(Y,S,C|\nu,\Lambda_X)\log p(Y,S,C|\nu',\Lambda_X)$$

第２ステップは、最大化ステップ（Ｍステップ）と呼ぶ。この第２ステップで、Ｑ（ν′｜ν）を最大にする◆

$$\nu'=\arg\max_{\nu} Q(\nu'|\nu) \qquad (15)$$

である。Ｑ（ν′｜ν）≧Ｑ（ν｜ν）である場合、ｐ（Ｙ｜ν′，Λx）≧ｐ（Ｙ｜ν，Λx）となることを示すことができる。従って、式１４および１５のＥステップおよびＭステップを反復して適用した場合に尤度が非減少であることが保証される。反復は、尤度の増大がある所定のしきい値未満になるまで継続される。

＊【００５６】一般に、式２の関数Ｆν（・）はＹのブロックを異なるサイズのＸのブロックに写像することができる。しかし、簡単のため、この関数は、次式のように、Ｙの各フレームを対応するＸのフレームに写像するようなものであると仮定する。

$$x_t=f_{\nu}(y_t) \qquad (16)$$

上記のような連続密度ＨＭＭによって与えられるΛxに

18

＊であると仮定する。

【数１０】

※られる正規分布である。

【数１１】

★＝｛ｃ１，ｃ２，...，ｃＴ｝を、すべてのミクスチャ成分列の集合とする。すると、式８は次のように書くことができる。

【数１２】

【数１３】

◆ν′の値を見つける。すなわち、

【数１５】

【数１６】

より、この補助関数は次のように書き換えることができ

19 20

る。

【数１７】：

$$Q(\nu'|\nu)=\sum_{S}\sum_{C}p(Y,S,C|\nu,\Lambda_X)\log\prod_{t=1}^{T}a_{s_{t-1},s_t}w_{s_t,c_t}p_y(y_t|s_t,c_t,\nu',\Lambda_X) \tag{17}$$

ただし、ｐｙ（ｙｔ｜ｓｔ，ｃｔ，ν，Λｘ）は、ランダム変数ｙｔの確率密度関数である。これは、式１１によって与えられるランダム変数ｘｔの密度関数と関係ｘｔ＝ｆ＊ν（ｙｔ）から導出することができる。ｙｔの密度は次のように書くことができる。

【数１８】

$$p_y(y_t|s_t,c_t,\nu,\Lambda_X)=\frac{N\left(f_\nu(y_t);\mu_{s_t,c_t},C_{s_t,c_t}\right)}{|J_\nu(y_t)|} \tag{18}$$

ただし、Ｊν（ｙｔ）は、（ｉ，ｊ）成分が次式によって与えられるようなヤコビ行列である。

※【数１９】

$$J_{\nu i,j}=\frac{\partial y_{t,i}}{\partial f_{\nu,j}(y_t)} \tag{19}$$

ただし、ｙｔ，ｉはｙｔの第ｉ成分であり、ｆν，ｊ（ｙｔ）はｆν（ｙｔ）の第ｊ成分である。さらに、式１７は次★のように書き換えることができる。

【数２０】

$$Q(\nu'|\nu)=\sum_{S}\sum_{C}p(Y,S,C|\nu,\Lambda_X)\cdot\sum_{t=1}^{T}\left\{\log a_{s_{t-1},s_t}+\log w_{s_t,c_t}+\log\frac{N\left(f_{\nu'}(y_t);\mu_{s_t,c_t},C_{s_t,c_t}\right)}{|J_{\nu'}(y_t)|}\right\} \tag{20}$$

これはさらに次式のように書ける。

【数２１】

$$\begin{aligned}Q(\nu'|\nu)=&\sum_{n=1}^{N}p(Y,s_1=n|\nu,\Lambda_X)\log a_{s_0,n}\\&+\sum_{t=2}^{T}\sum_{n=1}^{N}\sum_{l=1}^{N}p(Y,s_t=n,s_{t-1}=l|\nu,\Lambda_X)\log a_{l,n}\\&+\sum_{t=1}^{T}\sum_{n=1}^{N}\sum_{m=1}^{M}p(Y,s_t=n,c_t=m|\nu,\Lambda_X)\log w_{n,m}\\&+\sum_{t=1}^{T}\sum_{n=1}^{N}\sum_{m=1}^{M}p(Y,s_t=n,c_t=m|\nu,\Lambda_X)\log N\left(f_{\nu'}(y_t);\mu_{n,m},C_{n,m}\right)\\&-\sum_{t=1}^{T}\sum_{n=1}^{N}\sum_{m=1}^{M}p(Y,s_t=n,c_t=m|\nu,\Lambda_X)\log|J_{\nu'}(y_t)|\end{aligned} \tag{21}$$

ここで、ａｓ０，ｎは状態ｎの初期確率である。式２１の補助関数を計算する際には、ν′を含む項にのみ興味がある。従って、式１１を用いて、この補助関数を次の☆ように書くことができる。

【数２２】

$$Q(\nu'|\nu)=\sum_{t=1}^{T}\sum_{n=1}^{N}\sum_{m=1}^{M}\gamma_t(n,m)\cdot\left[-\frac{1}{2}\left(f_{\nu'}(y_t)-\mu_{m,n}\right)^T C_{n,m}^{-1}\left(f_{\nu'}(y_t)-\mu_{m,n}\right)-\log|J_{\nu'}(y_t)|\right] \tag{22}$$

ただし、γｔ（ｎ，ｍ）＝ｐ（Ｙ，ｓｔ＝ｎ，ｃｔ＝ｍ｜ν，Λｘ）は、Ｙと、観測値ｙｔを生成する状態ｎからのミクスチャｍとの結合尤度である。次のように、前方後方アルゴリズムを使用して確率γｔ（ｎ，ｍ）を計算す

ることができる。 【数２３】

$$\gamma_t(n,m)=\alpha_t(n)\beta_t(n)\frac{w_{n,m}N\left(f_v(y_t);\mu_{n,m},C_{n,m}\right)}{\sum_{j=1}^{M}w_{n,j}N\left(f_v(y_t);\mu_{n,j},C_{n,j}\right)} \quad (23)$$

ただし、 【数２４】

$$\alpha_t(n)=p\left(y_1,y_2,\ldots,y_t,s_t=n|v,\Lambda_X\right) \quad (24)$$

【数２５】

$$\beta_t(n)=p\left(y_{t+1},y_{t+2},\ldots,y_T|v,s_t=n,\Lambda_X\right) \quad (25)$$

である。前方後方アルゴリズムを使用して、αｔ（ｎ）およびβｔ（ｎ）を反復して計算することができる。

【００５７】ν′に関するＱ（ν′｜ν）の最大値を見つけるために、勾配上昇アルゴリズムのような任意の山登り法を使用することができる。しかし、場合によって＊は、式２２の右辺をν′で微分してその零点を解くことにより、陽に解を導くことができる。すなわち、次のようなν′を見つけることができる。

【数２６】

$$\frac{\partial}{\partial v'}\sum_{t=1}^{T}\sum_{n=1}^{N}\sum_{m=1}^{M}\gamma_t(n,m)\cdot\left[-\frac{1}{2}\left(f_{v'}(y_t)-\mu_{m,n}\right)^T C_{n,m}^{-1}\left(f_{v'}(y_t)-\mu_{m,n}\right)-\log\left|J_{v'}(y_t)\right|\right]=0 \quad (26)$$

【００５８】式８の尤度の最大にするＥＭアルゴリズムは、式２２からＱ（ν′｜ν）を計算（Ｅステップ）した後、式２６からν′を見つける（Ｍステップ）ことによって実行される。その後、この値を式２２のνに代入し、アルゴリズムは反復して実行される。

【００５９】分節ｋ平均アルゴリズムとのアナロジー ※で、分節ＭＬアプローチを使用して、式１２の尤度ｐ（Ｙ｜ν，Λx）を直接最大にする代わりに、観測値と状態列の結合尤度ｐ（Ｙ，Ｓ｜ν，Λx）を最大にすることも可能である。この場合、反復推定手続きは次のようになる。

【数２７】

$$S^l=\arg\max_S p\left(Y,S|v^l,\Lambda_X\right) \quad (27)$$

【数２８】

$$v^{l+1}=\arg\max_v p\left(Y,S^l|v,\Lambda_X\right) \quad (28)$$

こうして、まず最尤状態列Ｓ^{l}を見つけた後、この状態列の条件付きで発声Ｙの尤度を最大にするν^{l+1}を見つける。ビタビアルゴリズムを使用して、最適な状態列Ｓ^{l}を見つけることが可能であり、ＥＭアルゴリズムを使★用してν^{l+1}を見つけることが可能である。容易に示されるように、上記のＥＭ手続きは、γｔ（ｎ，ｍ）が次式によって定義されることを除いてはやはり成り立つ。

【数２９】

$$\gamma_t(n,m)=\begin{cases}\dfrac{w_{n,m}N\left(f_v(y_t);\mu_{n,m},C_{n,m}\right)}{\sum_{j=1}^{M}w_{n,j}N\left(f_v(y_t);\mu_{n,j},C_{n,j}\right)} & s_t^l=n\text{の場合}\\ 0 & \text{それ以外の場合}\end{cases} \quad (29)$$

【００６０】簡単のため、ｆν（ｙｔ）は各成分に別個に作用し（すなわち、ｘｔ，ｉ＝ｆν，ｉ（ｙｔ，ｉ））、共分散行列Ｃｎ，ｍは対角形である（すなわち、Ｃｎ，ｍ＝ｄｉａｇ（σ^{2}ｎ，ｍ））と仮定する。以下では、表現を容易にするため、ベクトルの第ｉ成分を表す添字ｉの参照を省略する。次の形の関数を考える。

【数３０】

$$f_v(y_t) = ag(y_t) + b \qquad (30)$$

ただし、ｇ（ｙｔ）はｙｔについての既知の（おそらくは非線形の）微分可能関数であり、ｖ＝｛ａ，ｂ｝は既知パラメータの集合である。すると、式２２の補助関数は＊次のように書くことができる。

【数３１】

$$Q(a',b'|a,b) = \sum_{t=1}^{T}\sum_{n=1}^{N}\sum_{m=1}^{M}\gamma_t(n,m)\left[-\frac{1}{2}\frac{\left(a'g(y_t)+b'-\mu_{m,n}\right)^2}{\sigma_{n,m}^2} + \log a'\right] \qquad (31)$$

式３１をａ′およびｂ′のそれぞれに関する微分をとって０とおくことにより、※

【数３２】

$$\sum_{t=1}^{T}\sum_{n=1}^{N}\sum_{m=1}^{M}\gamma_t(n,m)\left[\frac{1}{a'}-\frac{\left(a'g(y_t)+b'-\mu_{m,n}\right)g(y_t)}{\sigma_{n,m}^2}\right]=0 \qquad (32)$$

および

【数３３】

$$\sum_{t=1}^{T}\sum_{n=1}^{N}\sum_{m=1}^{M}\gamma_t(n,m)\frac{a'g(y_t)+b'-\mu_{m,n}}{\sigma_{n,m}^2}=0 \qquad (33)$$

を得る。式３２および３３は、推定値ａ′およびｂ′について陽に解くことができる。★０の特別の場合、すなわち、

【数３４】

【００６１】次に、加法的バイアスｂｔに対応する式３

$$x_t = y_t - b_t \qquad (34)$$

の場合を考える。各成分についてａ＝１、ｇ（ｙｔ）＝ｙｔ、およびｂ＝－ｂｔの場合、式３４は式３０と等価となる。観測値がスペクトラム領域にある場合、ｂｔは加法的ノイズスペクトラムと解釈することができる。一方、観測値がケプストラム領域すなわち対数エネルギー領域にある場合、ｂｔは例えばトランスデューサまたはチャネルによる畳込みフィルタ効果に対応する。

【００６２】バイアスｂｔは、ある発声に対して固定したものとして、または、時間とともに変動するものとして、モデル化することができる。時変バイアスの例に☆は、区分的一定バイアス、あるいは信号状態依存バイアスがある。あるいは、バイアスを確率論的にモデル化し、歪みをモデル空間（詳細は第４節）で見ることも可能である。この節では、状態依存バイアスおよび固定バイアスの場合について考察する。

【００６３】まず状態依存の場合を考える。バイアスはＨＭＭ状態ごとに変動する。各音声状態ｎに対応して特定のバイアス項ｂｎがあると仮定する。式２２の補助関数を次のように書くことができる。

【数３５】

$$Q(b'|b) = \sum_{t=1}^{T}\sum_{n=1}^{N}\sum_{m=1}^{M}\gamma_t(n,m)\left[-\sum_{i=1}^{D}\frac{\left(y_{t,i}-b'_{n,i}-\mu_{m,n,i}\right)^2}{2\sigma_{n,m,i}^2}\right] \qquad (35)$$

ただし、ｂ＝｛ｂ１，．．．，ｂｎ，．．．，ｂＮ｝である。式２６の再推定手続きには、式３５のｂ′ｎ，ｉに関する導関数を計算して０と等置することが必要である。その結◆果次式を得る。

【数３６】

$$\sum_{t=1}^{T}\sum_{m=1}^{M}\gamma_t(n,m)\frac{y_{t,i}-b'_{n,i}-\mu_{m,n,i}}{\sigma_{n,m,i}^2}=0 \qquad (36)$$

これは次式を与える。

【数３７】

$$b'_{n,j}=\frac{\sum_{t=1}^{T}\sum_{m=1}^{M}\gamma_t(n,m)\frac{y_{t,j}-\mu_{m,n,j}}{\sigma^2_{n,m,j}}}{\sum_{t=1}^{T}\sum_{m=1}^{M}\frac{\gamma_t(n,m)}{\sigma^2_{n,m,j}}} \quad (37)$$

単一の固定バイアスｂの場合、同様にして次式を示すことができる。

$$b'_i=\frac{\sum_{t=1}^{T}\sum_{n=1}^{N}\sum_{m=1}^{M}\gamma_t(n,m)\frac{y_{t,j}-\mu_{m,n,j}}{\sigma^2_{n,m,j}}}{\sum_{t=1}^{T}\sum_{n=1}^{N}\sum_{m=1}^{M}\frac{\gamma_t(n,m)}{\sigma^2_{n,m,i}}} \quad (38)$$

【００６４】式３７からわかるように、推定すべき状態依存バイアス項が多い場合に、小さいサンプルの効果によって推定の問題が起こることがある。しかし、状況によっては、状態依存バイアスには意味がある。例えば、線形フィルタリングに対する加法的ケプストラムバイアスは信号対ノイズ比（ＳＮＲ）が高い場合にのみ妥当である。ＳＮＲが低いときには、ノイズが優勢となり、チャネルフィルタリングに対する加法的バイアスモデルは不正確となる。これを扱う１つの方法は、そのバイアスがＳＮＲ依存であると仮定し、異なるＳＮＲ範囲に応じて異なるバイアスを推定することである。このようなアプローチを実現したものの１つに、音声および背景の分節に異なるバイアスを推定するものがある。これは、不一致の一部が電話チャネルによって引き起こされるような場合に有用であることがわかっている。これはおそらく電話チャネルに存在する加法的ノイズによるものである。その結果の詳細は第５節で説明する。

【００６５】４．モデル空間変換Ｇηの推定

前節では、歪みのある音声が原音声の定関数であると仮定した。この節では、歪みはランダムであるとみなし、それをモデル空間で見る。すなわち、歪みは原モデルΛｘを歪みのあるモデルΛｙに変換する（図４）。

【００６６】観測値列Ｙ＝｛ｙ１，...，ｙＴ｝が、原発声Ｘ＝｛ｘ１，...，ｘＴ｝および歪み列Ｂ＝｛ｂ１，...，ｂＴ｝と次式によって関係しているとする。

【数３９】

$$y_t=f(x_t,b_t) \quad (39)$$

すると、ｘｔとｂｔが独立である場合、ｙｔの確率密度関数（ｐｄｆ）を次のように書くことができる。

【数４０】

$$p(y_t)=\iint_{H_t}p(x_t)p(b_t)dx_t db_t \quad (40)$$

ただし、Ｈｔは式３９によって与えられる曲線である。

【００６７】前のように、ＸはＨＭＭの集合Λｘによってモデル化される。Ｂの統計モデルがΛＢによって与えられるとする。ΛＢはＨＭＭまたは混合ガウス密度とすることが可能である。この考察では、ΛＢは、次のような、対角形共分散行列を有する単一のガウス密度であると仮定する。

【数４１】

$$p(b_t)=N(b_t;\mu_b,\sigma_b^2) \quad (41)$$

さらに、式３４のときのように、加法的バイアスという特別の場合を考え、曲線Ｈｔが次式によって与えられるとする。

【数４２】

$$y_t=x_t+b_t \quad (42)$$

これらの仮定のもとで、Λｙの構造はΛｘの構造と同じく保たれる。Λｙの各ミクスチャ成分の平均および分散は、次のように、平均μｂおよび分散σｂ²を、Λｘにおける対応するミクスチャ成分の平均および分散に加えることによって導出される。

【数４３】

$$\mu_y=\mu_x+\mu_b \quad (43)$$

27 28

【数４４】

$$\sigma_y^2 = \sigma_x^2 + \sigma_b^2 \tag{44}$$

式４３および４４は、式３のモデル変換Ｇη（・）を定義し、パラメータηはμbおよびσb²によって与えられる。ΛBがHMMや混合ガウス密度のようにさらに複雑である場合、ΛYの構造は、状態およびミクスチャ成分の数が異なるという点で、ΛXの構造とは異なることも＊ある。

【００６８】ここで、式１３のパラメータ推定問題を次のように書くことができる。

【数４５】

$$\eta' = \left(\mu_b', \sigma_b^{2'}\right) = \arg\max_{\mu_b, \sigma_b^2} \sum_S \sum_C p\left(Y, S, C | \eta, \Lambda_X\right) \tag{45}$$

$$= \arg\max_{\mu_b, \sigma_b^2} \sum_S \sum_C \prod_{t=1}^{T} a_{s_{t-1}, s_t} w_{s_t, c_t} N\left(y_t; \mu_{s_t, c_t} + \mu_b, \sigma_{s_t, c_t}^2 + \sigma_b^2\right)$$

再び、ＥＭアルゴリズムを使用して反復してμbおよびσb²を推定することができる。容易に示されるように、式２１に対応する補助関数は次のように書くことができ※る。

【数４６】

$$Q(\eta'|\eta) = \tag{46}$$

$$-\sum_{t=1}^{T}\sum_{n=1}^{N}\sum_{m=1}^{M}\gamma_t(n,m)\sum_{i=1}^{D}\left[\frac{1}{2}\log\left(\sigma_{n,m,i}^2 + \sigma_{b_i}^{2'}\right) + \frac{\left(y_{t,i} - \mu_{b_i}' - \mu_{m,n,i}\right)^2}{2\left(\sigma_{n,m,i}^2 + \sigma_{b_i}^{2'}\right)}\right]$$

この関数をη′＝（μb′，σb²′）に関して最大化することによって、σb²′に対する閉じた表式は得られないが、式３８と類似の表式がμb′について次のように★得られる。

【数４７】

$$\mu_{b_i}' = \frac{\sum_{t=1}^{T}\sum_{n=1}^{N}\sum_{m=1}^{M}\gamma_t(n,m)\dfrac{y_{t,i} - \mu_{m,n,i}}{\sigma_{n,m,i}^2 + \sigma_{b_i}^{2'}}}{\sum_{t=1}^{T}\sum_{n=1}^{N}\sum_{m=1}^{M}\dfrac{\gamma_t(n,m)}{\sigma_{n,m,i}^2 + \sigma_{b_i}^{2'}}} \tag{47}$$

【００６９】σb²′を推定する問題への１つのアプローチは、分散σb²が信号状態依存であり、信号分散と次式☆☆によって関係していると仮定することである。

【数４８】

$$\sigma_{b,n,m,i}^2 = \alpha_i \sigma_{n,m,i}^2 \tag{48}$$

ただし、αiは分散の第ｉ成分に対するスケールファクタである。式４８を式４６に代入し、αiに関して最大◆40◆化することによって、次式を得る。

【数４９】

$$1 + \alpha_i = \frac{1}{T}\sum_{t=1}^{T}\sum_{n=1}^{N}\sum_{m=1}^{M}\gamma_t(n,m)\frac{\left(y_{t,i} - \mu_{m,n,i} - \mu_{b_i}'\right)^2}{\sigma_{n,m,i}^2} \tag{49}$$

式４２の場合のような物理的意味づけを式４８の仮定に対してすることは容易にはできないが、これにより式４７および４９に示されるようにパラメータμbおよびσb²の閉じた推定が得られる。αi＞－１は分散膨張（α＞０）および分散収縮（α＜０）の両方に対応する。

【００７０】式４２と整合する別のアプローチは、尤度ｐ（Ｙ｜η，ΛＸ）を次のように書くことである。

【数５０】

$$p(Y|\eta,\Lambda_X) = \sum_S \sum_C \iint_H p(X,B,S,C|\eta,\Lambda_X)\,dXdB \qquad (50)$$

ただし、∬Hは次式によって与えられるＴ重積分である。

$$\iint_H dXdB = \prod_{t=1}^{T} \iint_{H_t} dx_t db_t \qquad (51)$$

対応して、新たな補助関数を次のように定義することができる。

$$Q(\eta'|\eta) = \sum_S \sum_C \iint_H p(X,B,S,C|\eta,\Lambda_X) \log p(X,B,S,C|\eta',\Lambda_X)\,dXdB \qquad (52)$$

このアプローチは、ノイズのある観測値が与えられた場合に原音声モデルのパラメータの推定に対する一般的表式を導出するためにとられている。ただし、音声および歪みは両方とも混合ガウス密度によってモデル化される。この節で考察した問題は、歪みのある音声が与えられた場合に歪みのあるモデルのパラメータを見つける問★題の逆である。音声モデルΛxはHMMであり、加法的歪みは式４１のような単一のガウス密度としてモデル化される。これらの条件下で、微分により、次のような再推定公式が得られる。

【数５３】

$$\mu'_{b_i} = \frac{\sum_{t=1}^{T}\sum_{n=1}^{N}\sum_{m=1}^{M}\gamma_t(n,m)E(b_{t,j}|y_{t,j},s_t=n,c_t=m,\eta,\Lambda_X)}{\sum_{t=1}^{T}\sum_{n=1}^{N}\sum_{m=1}^{M}\gamma_t(n,m)} \qquad (53)$$

【数５４】

$$\sigma'^{2}_{b_i} = \frac{\sum_{t=1}^{T}\sum_{n=1}^{N}\sum_{m=1}^{M}\gamma_t(n,m)E(b_{t,j}|y_{t,j},s_t=n,c_t=m,\eta,\Lambda_X)}{\sum_{t=1}^{T}\sum_{n=1}^{N}\sum_{m=1}^{M}\gamma_t(n,m)} - \mu'^{2}_{b_i} \qquad (54)$$

前のように、γｔ＝ｐ（Ｙ，ｓｔ＝ｎ，ｃｔ＝ｍ｜η，ΛＸ）は、Ｙと、観測値ｙｔを生成した変換済みモデルΛＹ＝Ｇη（ΛＸ）におけるｎ番目の状態の第ｍミクス☆チャとの結合尤度である。式５３および式５４における条件付き期待値は次のように評価される。

【数５５】

$$E(b_{t,j}|y_{t,j},s_t=n,c_t=m,\eta,\Lambda_X) = \mu_{b_i} + \frac{\sigma^2_{b_i}}{\sigma^2_{n,m,i}+\sigma^2_{b_i}}(y_{t,j}-\mu_{n,m,i}-\mu_{b_i}) \qquad (55)$$

【数５６】

$$E(b^2_{t,j}|y_{t,j},s_t=n,c_t=m,\eta,\Lambda_X) = \frac{\sigma^2_{b_i}\sigma^2_{n,m,i}}{\sigma^2_{b_i}+\sigma^2_{n,m,i}} + \left\{E(b_{t,j}|y_{t,j},s_t=n,c_t=m,\eta,\Lambda_X)\right\}^2 \qquad (56)$$

【００７１】式５５を調べると、ＥＭアルゴリズムの収束性について観察することができる。σbi²が小さい場合、収束は遅い。これは、われわれの実験（第５節）の場合もそうであり、異なるトランスデューサおよび伝送チャネルによる不一致の分散が小さいためである。決定論的バイアス（σbi²＝０）の極限の場合、推定値は全く変化しない。これは、式４７を使用してμbを推定し式５４を使用してσb²を推定することによって補正する

31

ことができる。

【００７２】式３４の加法的モデルのもとで特徴空間およびモデル空間におけるバイアスパラメータを推定する方法を示した。しかし、加法的バイアスモデルはケプストラム特徴にのみ適用されている。われわれの実験では、ケプストラム特徴に加えて、デルタおよびデルターデルタケプストラム特徴ならびにデルタおよびデルターデルタ対数エネルギー特徴を使用した。確率論的マッチングアルゴリズムでは、デルタおよびデルターデルタ対数エネルギー特徴は変換しない。しかし、デルタケプストラムおよびデルターデルタケプストラムに対する不一致の効果は考慮する。特徴空間バイアスモデルでは、デ＊

$$\Delta C_{l,m} = \sum_{k=-K}^{K} Gk C_{l-k,m}$$

ただし、$\Delta C_{l,m}$および$C_{l,m}$はそれぞれ、ｌ番目の時間フレームに対するｍ番目のデルタケプストラム係数およびｍ番目のケプストラム係数である。Ｇは０．３７５に※

$$\Delta^2 C_{l,m} = \sum_{n=-N}^{N} Gn \Delta C_{l-n,m}$$

ただし、$\Delta^2 C_{l,m}$はｌ番目の時間フレームに対するｍ番目のデルターデルタケプストラム係数である。Ｇ＝０．３７５およびＮ＝１と選ぶ。異なるフレームに対するケプストラム係数は独立であると仮定すると、デルタケプ★

$$\sigma^2_{\Delta C_{l,m}} = \sum_{k=-K}^{K} G^2 k^2 \sigma^2_{C_{l-k,m}}$$

ただし、$\sigma^2 \Delta c_{l,m}$および$\sigma^2 c_{l,m}$は、ｌ番目の時間フレームのデルタケプストラムおよびケプストラムの第ｍ成分の分散である。同様に、デルターデルタケプストラム☆

$$\sigma^2_{\Delta^2 C_{l,m}} = \sum_{n=-N}^{N} \sum_{k=-K}^{K} G^4 n^2 k^2 \sigma^2_{C_{l-k-n,m}}$$

デルタおよびデルターデルタバイアス項の分散を推定することに興味がある。バイアスは分散がσ_b^2のｉ．ｉ．ｄ．ガウシアンであると仮定されるので、デルタバイア◆

$$\sigma^2_{\Delta b_i} = \sum_{k=-K}^{K} G^2 k^2 \sigma^2_{b_i}$$

同様に、デルターデルタバイアスの第ｉ成分の分散は次のように推定することができる。＊

$$\sigma^2_{\Delta^2 b_i} = \sum_{n=-N}^{N} \sum_{k=-K}^{K} G^4 n^2 k^2 \sigma^2_{b_i}$$

【００７３】観察されるように、歪みに使用される統計モデルは単純なガウス密度である。上記と同じ確率論的マッチングアルゴリズムは、より一般的なモデル変換の場合にも適用可能である。

【００７４】

32

＊ルタおよびデルターデルタケプストラム特徴は不一致によって影響を受けないと仮定する。すなわち、デルタおよびデルターデルタバイアスベクトルは０であると仮定する。これは、ケプストラムのバイアスが発声全体に対して一定であると仮定する場合には意味のある仮定である。同様に、モデル空間では、デルタおよびデルターデルタ平均値ベクトルが０であると仮定する。しかし、デルタおよびデルターデルタ分散については仮定しない。これらの分散ベクトルは以下のように推定する。デルタケプストラムは次式に従って計算される。

【数５７】

(57)

※固定された利得項であり、Ｋ＝２である。デルターデルタケプストラムは次式に従って計算される。

【数５８】

(58)

★ストラムの分散は、ケプストラムの分散を用いて次のように書くことができる。

【数５９】

(59)

☆の分散を次のように導出することができる。

【数６０】

(60)

◆スの第ｉ成分の分散は式５９を使用して次のように推定することができる。

【数６１】

(61)

＊【数６２】

(62)

【発明の効果】以上述べたごとく、本発明によれば、本発明に従って形成したシステムは、個々のアプリケーションごとにトレーニングする必要はないという効果がある。また、本発明に従って形成したシステムは、プールモデルやコードブックを使用する既知のシステムよりも

33

高い精度を有するという効果がある。

【図面の簡単な説明】

【図１】不一致問題を解決する既知のシステムの図である。

【図２】本発明に従って形成したシステムの第１実施例の図である。

【図３】本発明に従って形成したシステムの第２実施例の図である。

【図４】トレーニングと試験の不一致を示す図である。

【図５】式７の結合最大化を示す図である。

【図６】式７の結合最大化を示す図である。

【符号の説明】

１００　変換器

34

１０２　オフセット計算器

１０４　コードブック

１０６　比較器

１０８　HMM記憶装置

２００　変換器

２０２　変換プロセッサ

２０４　変換記憶装置

２０６　比較器

２０８　HMM記憶装置

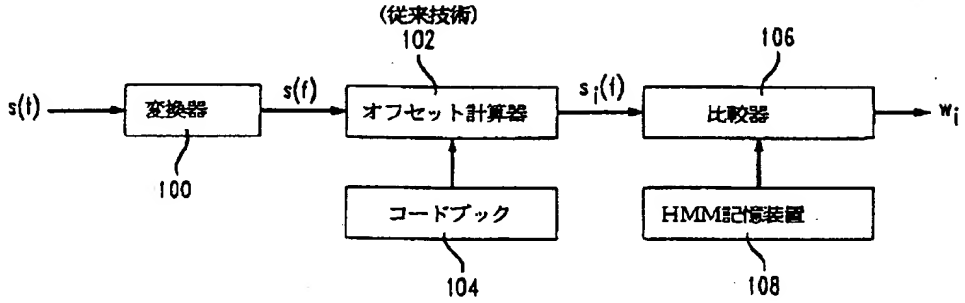
３００　変換器

３０２　比較器

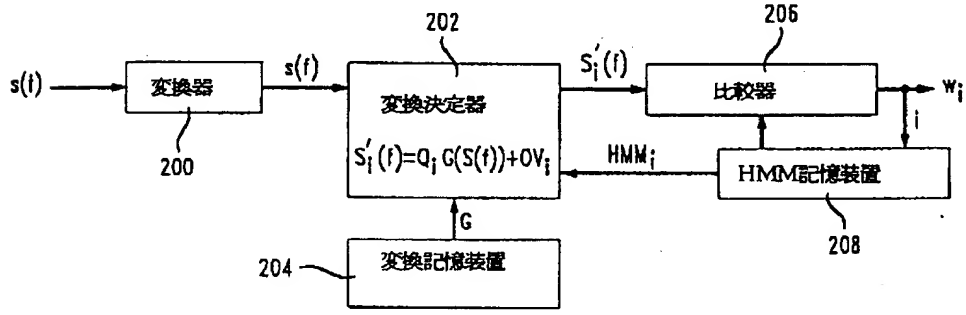
３０４　HMM記憶装置

３０６　変換プロセッサ

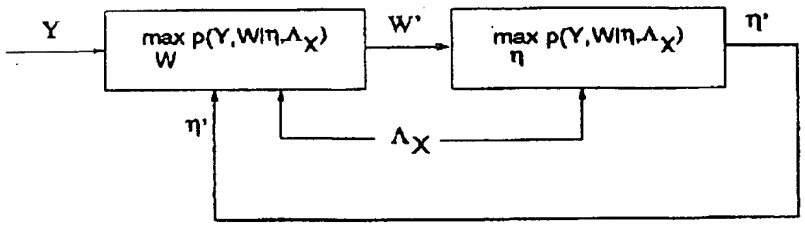
【図１】



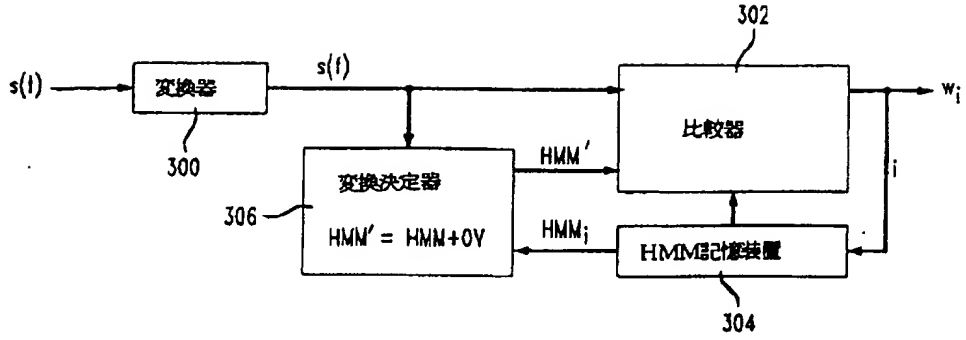
【図２】



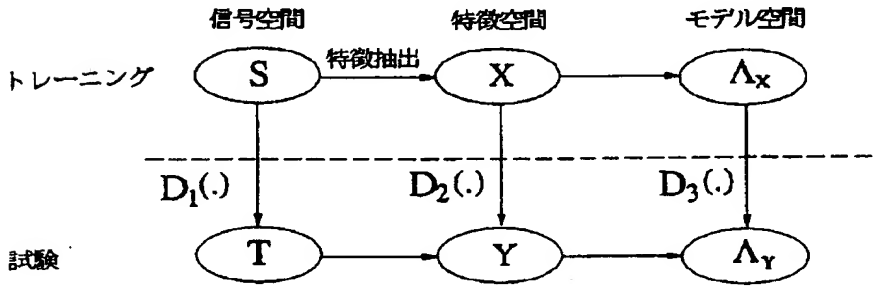
【図６】



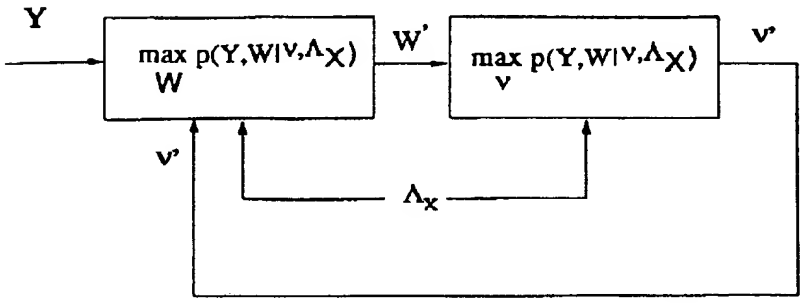
【図３】



【図４】



【図５】



フロントページの続き

(72)発明者　アナンス　サンカー
アメリカ合衆国，94555　カリフォルニア，
フレモント，トゥペロ　ストリート
34367